

SISTEM PREDIKSI RISIKO KETERLAMBATAN DISTRIBUSI PANGAN PROGRAM MAKANAN BERGIZI GRATIS BERBASIS MACHINE LEARNING

Ismail¹, Nur Fadillah Amiruddin^{2*}, Hasna³, Zinta⁴, Kamis Tati⁵

^{1,2,3,4,5} Universitas Lamappapoleonro, Jl. Salotungo No.62, Watansoppeng, Soppeng, Sulawesi Selatan

¹ismail@unipol.ac.id, ^{2*}nurfadillahamiruddin7@gmail.com, ³hasnaenna2005@gmail.com,

⁴zintaaa0101@gmail.com, ⁵kamistati239@gmail.com

DOI: <https://doi.org/10.46880/mtk.v12i2.5930>

ABSTRACT

Food distribution in the Free Lunch Program (MBG) requires timeliness to maintain food quality and service effectiveness to beneficiaries. Delays can be influenced by distribution distance, number of recipients, weather, road conditions, delivery time, and vehicle type. This study aims to develop a machine learning-based prediction model for the risk of delays in MBG food distribution. The research method uses a quantitative approach with a dataset of 100 data samples from operational distribution scenarios in Soppeng Regency. Data were processed through cleaning, categorical variable coding, numeric variable normalization, 5-fold cross-validation splitting, model training, and performance evaluation. Four algorithms were compared: Random Forest, Decision Tree, K-Nearest Neighbor, and Logistic Regression. The test results showed that Random Forest achieved 92.00% accuracy, 92.00% precision, 92.00% recall, and 92.00% F1-score. Feature importance analysis showed that the number of recipients, distribution distance, and distribution time were the most dominant factors in determining the risk of delays. The proposed prediction system can be a tool for MBG distribution managers in identifying potential delays early and formulating more appropriate operational mitigation recommendations.

Keywords: *Distribution Delays, Machine Learning, Makanan Bergizi Gratis, Risk Prediction*

I. PENDAHULUAN

Program Makan Bergizi Gratis (MBG) menjadi salah satu program prioritas nasional yang bertujuan meningkatkan kualitas sumber daya manusia melalui pemenuhan kebutuhan gizi bagi kelompok sasaran. Mengacu pada Peraturan Presiden Nomor 83 Tahun 2024 tentang Badan Gizi Nasional, penerima manfaat program ini meliputi peserta didik pada jenjang PAUD, pendidikan dasar dan menengah, pendidikan keagamaan serta pesantren, anak berusia di bawah lima tahun, ibu hamil, dan ibu menyusui. Luasnya kelompok penerima tersebut memerlukan pengelolaan menu yang terencana, aman, dan disesuaikan dengan kebutuhan gizi masing-masing sasaran. Oleh karena itu, penyusunan menu MBG perlu mempertimbangkan ketersediaan bahan pangan lokal, kandungan energi dan protein, kecukupan zat gizi mikro, kualitas pangan, keberagaman menu, serta kemampuan biaya untuk setiap porsi [1].

Pelaksanaan Program Makan Bergizi Gratis melibatkan serangkaian kegiatan yang saling berhubungan, mulai dari perencanaan menu, pengadaan bahan pangan, pengolahan makanan, pemorsian, pengemasan, hingga pendistribusian kepada sekolah atau kelompok penerima. Keberhasilan program tersebut tidak hanya ditentukan oleh kandungan gizi dan kualitas menu yang disediakan, tetapi juga dipengaruhi oleh kelancaran sistem distribusi pangan. Makanan yang telah selesai diolah harus disalurkan secara tepat waktu, aman, dan tetap berada dalam kondisi layak konsumsi ketika diterima. Dengan demikian, sistem distribusi menjadi salah satu bagian penting yang menentukan keberhasilan pelaksanaan Program Makan Bergizi Gratis secara keseluruhan. Penelitian Kaylla dkk. menunjukkan bahwa efektivitas Program MBG tidak hanya bergantung pada kualitas menu, tetapi juga pada efektivitas sistem distribusi pasokan bahan pangan [2].

Dalam pelaksanaannya, proses distribusi pangan Program Makan Bergizi Gratis masih dapat menghadapi berbagai kendala operasional. Keterlambatan dapat terjadi pada tahap persiapan bahan baku, pengolahan, pemorsian, pengemasan, maupun saat makanan diantarkan ke lokasi penerima. Selain itu, jarak pengiriman, kondisi jalan, cuaca, jumlah penerima, waktu tempuh, dan kesiapan kendaraan dapat memengaruhi kelancaran distribusi. Kaylla dkk. menemukan bahwa distribusi MBG di wilayah Sariajadi masih mengalami kendala berupa keterlambatan pada tahap pemorsian akibat kesiapan bahan baku serta keterbatasan sarana transportasi. Penelitian tersebut juga menekankan perlunya peningkatan koordinasi internal dan fasilitas transportasi untuk menjaga ketepatan waktu, kualitas, dan keamanan pangan [2].

Permasalahan distribusi MBG juga tidak selalu disebabkan oleh faktor teknis, tetapi dapat berkaitan dengan lemahnya koordinasi dan komunikasi antarpetugas. Serang dkk. menemukan adanya keterlambatan distribusi secara rutin, koordinasi yang terfragmentasi, dan komunikasi yang masih bersifat reaktif dalam pelaksanaan MBG di salah satu sekolah di Kota Bitung. Kondisi tersebut menunjukkan bahwa masalah distribusi memiliki hubungan dengan tata kelola operasional dan koordinasi antaraktor yang terlibat. Oleh karena itu, diperlukan pengelolaan distribusi yang terencana, sistem komunikasi yang terpadu, dan pemanfaatan teknologi untuk membantu pengelola mengenali kemungkinan keterlambatan sebelum proses pengiriman dilakukan [3].

Keterlambatan distribusi pangan dapat menimbulkan berbagai dampak terhadap pelaksanaan program. Makanan yang diterima melewati waktu yang telah ditentukan berpotensi mengalami penurunan kualitas, kesegaran, dan keamanan untuk dikonsumsi.



Keterlambatan juga dapat mengganggu jadwal makan peserta didik dan menyebabkan tujuan program tidak tercapai secara maksimal. Apabila masalah tersebut terjadi secara berulang, tingkat kepercayaan penerima terhadap pelaksanaan program juga dapat menurun. Oleh karena itu, risiko keterlambatan perlu diketahui sedini mungkin agar pengelola dapat menentukan tindakan pencegahan, seperti mempercepat persiapan makanan, mengganti kendaraan, menyesuaikan waktu keberangkatan, atau memilih jalur distribusi yang lebih baik. Penataan manajemen rantai pasok dan penggunaan sistem pendukung dapat membantu meningkatkan ketepatan waktu serta keberhasilan pelaksanaan program [2], [3].

Perkembangan teknologi informasi memberikan peluang untuk mengolah data distribusi menjadi informasi yang dapat digunakan dalam pengambilan keputusan. Salah satu teknologi yang dapat digunakan adalah *machine learning*. Teknologi tersebut memungkinkan komputer mempelajari pola dari data distribusi sebelumnya dan menggunakan pola tersebut untuk memprediksi kemungkinan terjadinya keterlambatan pada pengiriman berikutnya. Dalam sistem prediksi, data seperti jarak pengiriman, jumlah penerima, waktu tempuh, kondisi jalan, cuaca, dan jenis kendaraan dapat dijadikan variabel masukan. Selanjutnya, sistem mengolah data tersebut untuk menghasilkan kategori risiko berupa tepat waktu, terlambat, atau sangat terlambat. Dengan adanya hasil prediksi, pengelola dapat memperoleh peringatan awal dan mengambil langkah pencegahan sebelum proses distribusi dilakukan.

Beberapa penelitian terdahulu telah membuktikan bahwa metode *machine learning* dapat diterapkan untuk memprediksi status dan risiko keterlambatan pengiriman. Arayadiba dkk. menerapkan algoritma *Decision Tree* untuk mengklasifikasikan risiko keterlambatan pengiriman barang dan memperoleh akurasi keseluruhan sebesar 71,01% [4]. Pambudi dkk. membandingkan metode *Logistic Regression*, *Random Forest*, dan *Artificial Neural Network* untuk memprediksi status pengiriman barang. Hasil penelitian tersebut menunjukkan bahwa *Random Forest* memperoleh akurasi terbaik sebesar 76,6%, sedangkan *Artificial Neural Network* memperoleh 73,81% dan *Logistic Regression* memperoleh 72,84% [5]. Sementara itu, Eriya dkk. membandingkan sembilan algoritma klasifikasi dalam memprediksi keterlambatan pengiriman pada rantai pasok. Hasil pengujian menunjukkan bahwa *Decision Tree* menjadi model terbaik pada data yang digunakan, dengan nilai F1 sebesar 99,41% untuk prediksi keterlambatan pengiriman [6]. Hasil tersebut menunjukkan bahwa *machine learning* dapat digunakan untuk mengidentifikasi risiko keterlambatan dan membantu pengambilan keputusan berbasis data.

Meskipun penelitian mengenai pelaksanaan dan distribusi Program Makan Bergizi Gratis telah dilakukan, penelitian tersebut lebih banyak membahas alur distribusi, kendala logistik, koordinasi, keamanan pangan, dan efektivitas pelaksanaan menggunakan pendekatan deskriptif atau kualitatif. Di sisi lain, penelitian tentang prediksi keterlambatan berbasis

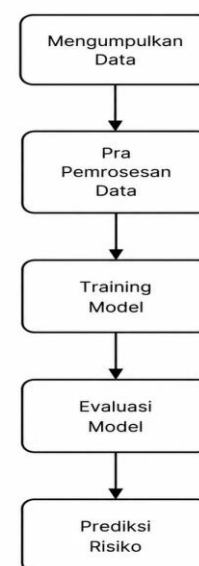
machine learning lebih banyak diterapkan pada pengiriman barang, perusahaan ekspedisi, dan rantai pasok secara umum. Berdasarkan kondisi tersebut, penerapan *machine learning* secara khusus untuk memprediksi risiko keterlambatan distribusi pangan pada Program Makan Bergizi Gratis masih perlu dikembangkan. Penelitian ini memiliki perbedaan karena menggabungkan permasalahan distribusi pangan MBG dengan teknologi prediksi berbasis data untuk menghasilkan informasi mengenai tingkat risiko keterlambatan.

Berdasarkan permasalahan tersebut, penelitian ini mengembangkan Sistem Prediksi Risiko Keterlambatan Distribusi Pangan Program Makanan Bergizi Gratis Berbasis Machine Learning. Variabel yang digunakan meliputi jarak pengiriman, jumlah penerima, waktu tempuh, kondisi jalan, kondisi cuaca, dan jenis kendaraan. Data distribusi diolah menggunakan algoritma *machine learning* untuk menghasilkan klasifikasi tepat waktu, terlambat, dan sangat terlambat. Penelitian ini bertujuan membangun sistem yang dapat membantu pengelola memperkirakan risiko keterlambatan distribusi pangan secara lebih awal. Sistem yang dikembangkan diharapkan dapat mendukung pengambilan keputusan, meningkatkan ketepatan waktu pengiriman, mengurangi risiko penurunan kualitas makanan, dan membantu meningkatkan efektivitas pelaksanaan Program Makan Bergizi Gratis.

II. METODE

Jenis dan Alur Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental. Data distribusi diolah menjadi dataset klasifikasi, kemudian dilakukan pelatihan dan pengujian model *machine learning*. Alur penelitian meliputi pengumpulan data, prapemrosesan, pembagian validasi, pelatihan algoritma, evaluasi model, dan perumusan rekomendasi risiko. Alur tersebut ditunjukkan pada Gambar 1.



Gambar 1. Metodologi Penelitian

Dataset dan Variabel

Dataset yang digunakan terdiri atas 100 data sampel skenario operasional distribusi pangan MBG di Kabupaten Soppeng. Setiap data merepresentasikan satu aktivitas distribusi menuju satu titik penerima manfaat. Variabel input meliputi jarak distribusi, jumlah penerima, cuaca, kondisi jalan, waktu distribusi, dan jenis kendaraan. Variabel target adalah status keterlambatan yang dibagi menjadi tiga kelas.

Tabel 1. Variabel Penelitian

No	Variabel	Tipe	Keterangan
1	Jarak	Numerik	Jarak pengiriman dalam kilometer
2	Jumlah penerima	Numerik	Jumlah penerima manfaat pada lokasi tujuan
3	Cuaca	Kategorikal	Cerah, berawan, hujan, hujan lebat
4	Kondisi jalan	Kategorikal	Baik, sedang, rusak ringan, rusak berat
5	Waktu distribusi	Kategorikal	Pagi, siang, sore
6	Kendaraan	Kategorikal	Mobil box, pickup, motor roda tiga
7	Status	Kategorikal	Tepat Waktu, Terlambat, Sangat Terlambat

Prapemrosesan Data

Tahap prapemrosesan dilakukan untuk memastikan data siap digunakan dalam proses pelatihan. Variabel numerik dinormalisasi menggunakan StandardScaler, sedangkan variabel kategorikal dikonversi menjadi bentuk numerik menggunakan one-hot encoding. Selanjutnya, model dievaluasi menggunakan stratified 5-fold cross validation agar proporsi setiap kelas tetap terjaga pada setiap lipatan data. Kategori risiko ditetapkan menjadi tiga kelas. Kelas Tepat Waktu menunjukkan pengiriman sesuai jadwal, kelas Terlambat menunjukkan adanya keterlambatan sedang, sedangkan kelas Sangat Terlambat menunjukkan kondisi risiko tinggi yang memerlukan prioritas penanganan.

Algoritma dan Evaluasi

Empat algoritma digunakan dalam penelitian ini, yaitu Random Forest, Decision Tree, K-Nearest Neighbor, dan Logistic Regression. Evaluasi dilakukan menggunakan akurasi, presisi, recall, dan F1-score. Akurasi digunakan untuk melihat jumlah prediksi benar secara keseluruhan. Presisi menggambarkan ketepatan model ketika memberikan suatu label, recall menggambarkan kemampuan model menemukan semua data pada kelas tertentu, sedangkan F1-score menunjukkan keseimbangan antara presisi dan recall.

$$F1\text{-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

III. HASIL DAN PEMBAHASAN

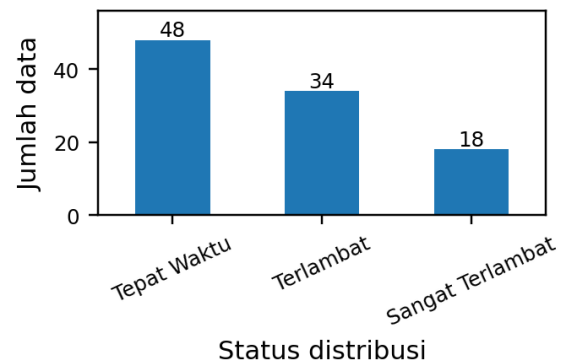
Karakteristik Dataset

Distribusi kelas pada dataset menunjukkan bahwa 48 data termasuk Tepat Waktu, 34 data termasuk

Terlambat, dan 18 data termasuk Sangat Terlambat. Komposisi ini menggambarkan bahwa kasus keterlambatan tetap cukup dominan untuk dianalisis, karena gabungan kelas Terlambat dan Sangat Terlambat mencapai 52% dari total data.

Tabel 2. Distribusi Kelas Status Distribusi

Status distribusi	Jumlah	Persentase
Tepat Waktu	48	48%
Terlambat	34	34%
Sangat Terlambat	18	18%
Total	100	100%



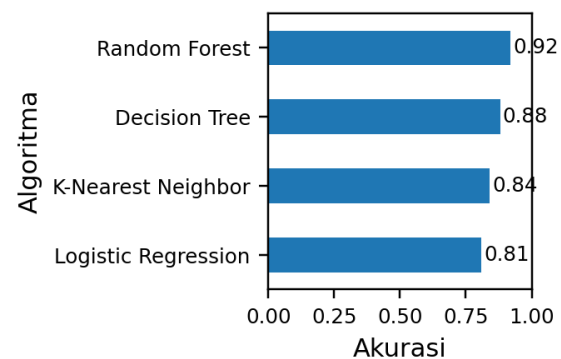
Gambar 2. Grafik Distribusi Status Distribusi

Pada Gambar 2 terlihat bahwa kasus Tepat Waktu merupakan kelas terbesar, namun jumlah keterlambatan tidak dapat diabaikan. Hal ini memperkuat kebutuhan sistem prediksi karena pengelola distribusi harus dapat membedakan kondisi normal, keterlambatan sedang, dan keterlambatan berat sebelum pengiriman dilakukan.

Hasil Perbandingan Algoritma

Tabel 3. Hasil Evaluasi Model Machine Learning

Algoritma	Akurasi	Presisi	Recall	F1-Score
Random Forest	0.92	0.92	0.92	0.92
Decision Tree	0.88	0.89	0.88	0.88
K-Nearest Neighbor	0.84	0.84	0.84	0.83
Logistic Regression	0.81	0.82	0.81	0.80



Gambar 3. Grafik Perbandingan Akurasi Algoritma

Berdasarkan Tabel 3 dan Gambar 3, Random Forest memperoleh akurasi 0,92, presisi 0,92, recall 0,92, dan F1-score 0,92. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan mayoritas status

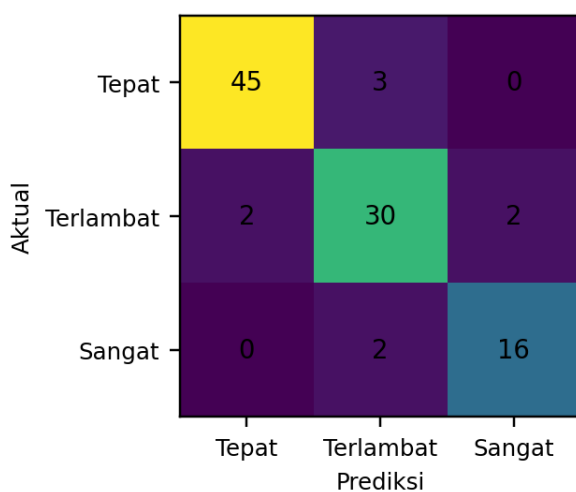
distribusi dengan baik. Decision Tree menghasilkan akurasi 0,88, sedangkan K-Nearest Neighbor dan Logistic Regression memperoleh akurasi 0,84 dan 0,81. Random Forest dipilih sebagai model utama karena memberikan performa tertinggi, stabilitas prediksi, kemampuan menangani variabel kategorikal hasil encoding, dan menyediakan informasi feature importance untuk interpretasi faktor keterlambatan.

Perbedaan kinerja antarmodel menunjukkan bahwa pola keterlambatan distribusi tidak sepenuhnya linear. Variabel seperti cuaca, kondisi jalan, dan waktu pengiriman dapat berinteraksi dengan jarak dan jumlah penerima. Model ensemble seperti Random Forest dinilai sesuai karena mampu membangun banyak pohon keputusan untuk mempelajari pola tersebut secara lebih fleksibel.

Confusion Matrix Random Forest

Tabel 4. Confusion Matrix Model Random Forest

Aktual / Prediksi	Tepat Waktu	Terlambat	Sangat Terlambat
Tepat Waktu	45	3	0
Terlambat	2	30	2
Sangat Terlambat	0	2	16



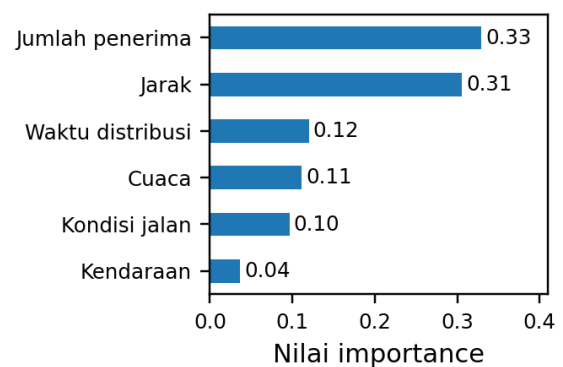
Gambar 4. Visualisasi Confusion Matrix Random Forest

Confusion matrix pada Tabel 4 menunjukkan bahwa dari 48 data aktual Tepat Waktu, 45 data diprediksi benar dan 3 data diprediksi sebagai Terlambat. Pada kelas Terlambat, 30 dari 34 data diprediksi benar, sedangkan 2 data masuk ke kelas Tepat Waktu dan 2 data masuk ke kelas Sangat Terlambat. Pada kelas Sangat Terlambat, 16 dari 18 data diprediksi benar. Kesalahan terbesar terjadi pada batas antara Terlambat dan Sangat Terlambat karena kedua kondisi tersebut memiliki karakteristik operasional yang berdekatan.

Analisis Feature Importance

Tabel 5. Faktor Dominan Risiko Keterlambatan

Variabel	Nilai importance	Interpretasi
Jumlah penerima	0.33	Paket lebih banyak menambah waktu persiapan dan distribusi.
Jarak	0.31	Semakin jauh lokasi tujuan, semakin besar peluang keterlambatan.
Waktu distribusi	0.12	Pengiriman siang/sore berisiko lebih tinggi dibanding pagi.
Cuaca	0.11	Hujan dan hujan lebat meningkatkan hambatan perjalanan.
Kondisi jalan	0.10	Jalan rusak memperlambat kendaraan distribusi.
Kendaraan	0.04	Kapasitas kendaraan memengaruhi efisiensi pengiriman.



Gambar 5. Grafik Feature Importance Random Forest

Hasil feature importance pada Tabel 5 dan Gambar 5 menunjukkan bahwa jumlah penerima memiliki pengaruh terbesar dengan nilai 0,33, diikuti jarak distribusi 0,31 dan waktu distribusi 0,12. Temuan ini menunjukkan bahwa keterlambatan paling banyak dipengaruhi oleh volume layanan dan beban jarak distribusi. Cuaca dan kondisi jalan tetap berpengaruh, terutama pada kondisi hujan lebat atau jalan rusak, tetapi nilainya lebih rendah dibanding jarak dan jumlah penerima pada dataset yang digunakan.

Dari sisi manajerial, hasil ini dapat diterjemahkan menjadi rekomendasi operasional. Lokasi dengan jarak jauh dan jumlah penerima besar perlu dijadwalkan lebih awal atau menggunakan kendaraan berkapasitas lebih sesuai. Pengiriman pada sore hari perlu dihindari untuk rute berisiko tinggi, sedangkan titik tujuan dengan cuaca buruk dan kondisi jalan rusak perlu diberikan buffer waktu tambahan.

Rancangan Output Sistem Prediksi

Sistem yang diusulkan menerima input berupa jarak, jumlah penerima, cuaca, kondisi jalan, waktu distribusi, dan jenis kendaraan. Setelah data dimasukkan, sistem

melakukan prapemrosesan sesuai format model dan menghasilkan kelas prediksi risiko. Output sistem tidak hanya menampilkan status, tetapi juga rekomendasi tindakan untuk membantu petugas distribusi.

Tabel 6. Keluaran dan Rekomendasi Sistem

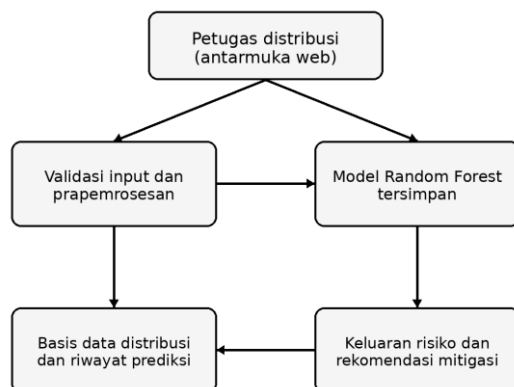
Kelas risiko	Keterangan	Rekomendasi tindakan
Tepat Waktu	Risiko rendah	Distribusi dilakukan sesuai jadwal standar.
Terlambat	Risiko sedang	Tambahkan buffer waktu dan cek kesiapan kendaraan.
Sangat Terlambat	Risiko tinggi	Prioritaskan rute, ubah jadwal, atau gunakan kendaraan alternatif.

Penerapan sistem prediksi ini dapat menjadi bagian dari sistem informasi distribusi pangan MBG. Pada tahap operasional, sistem dapat ditempatkan pada halaman prediksi yang terhubung dengan dataset distribusi dan laporan hasil pengiriman. Dengan demikian, pengelola dapat memantau tren keterlambatan, mengevaluasi faktor penyebab, dan memperbaiki perencanaan distribusi secara berkelanjutan.

Arsitektur Sistem Prediksi

Arsitektur sistem dirancang untuk memisahkan antarmuka pengguna, pengelolaan data, dan model machine learning. Petugas memasukkan informasi rencana distribusi melalui formulir, kemudian aplikasi memeriksa kelengkapan dan rentang nilai sebelum data diteruskan ke pipeline prapemrosesan. Pemisahan komponen tersebut memudahkan pembaruan model tanpa harus mengubah seluruh fungsi aplikasi.

Data kategorikal ditransformasikan menggunakan skema encoding yang sama dengan proses pelatihan, sedangkan data numerik diproses menggunakan parameter normalisasi yang telah tersimpan. Hasil transformasi dikirim ke model Random Forest untuk menghasilkan kelas risiko. Keluaran prediksi selanjutnya disimpan bersama waktu proses, nilai input, dan rekomendasi agar keputusan dapat ditelusuri kembali.



Gambar arsitektur konseptual sistem prediksi

Gambar 6. Arsitektur Konseptual Sistem Prediksi

Gambar 6 memperlihatkan hubungan antarkomponen sistem. Basis data tidak hanya digunakan untuk menyimpan dataset, tetapi juga menyimpan riwayat prediksi dan realisasi pengiriman. Riwayat tersebut penting untuk membandingkan prediksi dengan kejadian aktual serta menjadi sumber data ketika model diperbarui pada periode berikutnya.

Spesifikasi Kebutuhan Sistem

Kebutuhan fungsional disusun berdasarkan alur kerja petugas distribusi. Fungsi utama mencakup autentikasi, pengelolaan data, prediksi risiko, pemberian rekomendasi, pencatatan realisasi, serta penyusunan laporan. Rincian fungsi disajikan pada Tabel 7.

Tabel 7. Kebutuhan Fungsional Sistem

Kode	Fungsi	Keluaran
F1	Login dan pembatasan akses	Sesi pengguna
F2	Kelola data distribusi	Data tersimpan
F3	Proses prediksi risiko	Kelas risiko
F4	Tampilkan rekomendasi	Tindakan mitigasi
F5	Catat realisasi pengiriman	Riwayat aktual
F6	Cetak atau ekspor laporan	Laporan periodik

Setiap prediksi perlu terhubung dengan identitas data distribusi agar tidak terjadi duplikasi pencatatan. Pengguna juga perlu memperoleh pesan validasi ketika nilai jarak, jumlah penerima, atau pilihan kategori belum lengkap. Pada sisi laporan, sistem harus dapat menampilkan tanggal, tujuan, input utama, hasil prediksi, realisasi, dan tindakan yang dilakukan.

Selain fungsi utama, sistem memerlukan kebutuhan nonfungsional untuk menjaga kualitas layanan. Aspek tersebut meliputi kemudahan penggunaan, waktu respons, keamanan, keandalan penyimpanan, dan kemudahan pemeliharaan sebagaimana dirangkum pada Tabel 8.

Tabel 8. Kebutuhan Nonfungsional Sistem

Aspek	Kriteria
Kemudahan	Formulir singkat dan pesan mudah dipahami
Kinerja	Prediksi ditampilkan dalam beberapa detik
Keamanan	Akses menggunakan akun dan kata sandi
Keandalan	Data prediksi tersimpan dan dapat dilacak
Pemeliharaan	Model dapat diganti tanpa mengubah antarmuka
Portabilitas	Dapat diakses melalui peramban lokal

Pemenuhan kebutuhan nonfungsional tidak dinilai hanya dari tampilan aplikasi. Pengujian perlu memeriksa respons sistem saat data tidak lengkap, saat koneksi basis data terputus, dan saat model gagal dimuat. Dengan demikian, aplikasi tidak menampilkan hasil yang menyesatkan ketika terjadi gangguan teknis.

Mekanisme Inferensi dan Pengambilan Keputusan

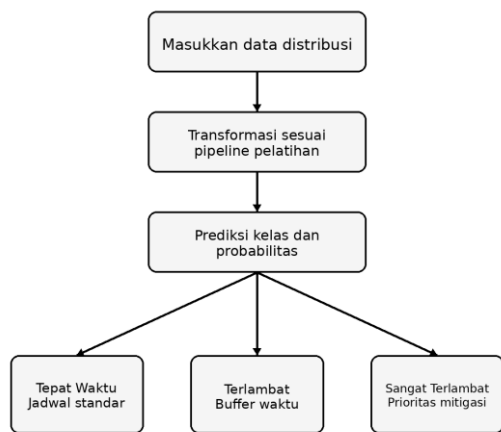
Proses inferensi dimulai setelah data input dinyatakan valid. Urutan transformasi harus sama dengan urutan pada proses pelatihan karena perubahan nama kolom, kategori, atau skala dapat menghasilkan prediksi yang tidak konsisten. Pipeline tersimpan digunakan untuk

mengurangi risiko perbedaan tersebut dan memastikan setiap data baru diproses dengan aturan yang sama [8].

Tabel 9. Tahapan Inferensi Data Baru

Tahap	Proses	Hasil
1	Validasi format input	Data layak proses
2	Encoding kategori	Vektor kategori
3	Normalisasi numerik	Nilai berskala
4	Prediksi Random Forest	Kelas risiko
5	Baca probabilitas kelas	Tingkat keyakinan
6	Pemetaan rekomendasi	Tindakan operasional

Apabila model menyediakan probabilitas kelas, nilai tersebut dapat digunakan sebagai informasi pendukung, bukan sebagai kepastian kejadian. Prediksi dengan selisih probabilitas yang kecil antarkelas perlu ditandai untuk ditinjau oleh petugas. Pendekatan ini menjaga agar keputusan penting tidak hanya bergantung pada keluaran otomatis, khususnya ketika kondisi lapangan berubah cepat.



Gambar 7. Alur Pengambilan Keputusan Prediksi

Gambar 7 menunjukkan bahwa setiap kelas risiko dihubungkan dengan tindakan yang berbeda. Kelas Tepat Waktu dapat mengikuti jadwal standar, kelas Terlambat memerlukan penambahan buffer waktu dan pemeriksaan kesiapan, sedangkan kelas Sangat Terlambat memerlukan prioritas rute atau kendaraan alternatif. Petugas tetap memiliki kewenangan untuk menyesuaikan rekomendasi berdasarkan informasi terbaru.

Skenario Operasional dan Mitigasi Risiko

Interpretasi prediksi perlu diterjemahkan menjadi tindakan yang dapat dilaksanakan. Satu kelas risiko dapat muncul dari kombinasi penyebab yang berbeda, sehingga rekomendasi tidak cukup hanya berdasarkan label akhir. Jarak jauh, jumlah penerima besar, hujan lebat, dan kondisi jalan rusak perlu ditinjau secara bersamaan ketika petugas menentukan prioritas distribusi.

Tabel 10. Contoh Skenario dan Mitigasi Operasional

Kondisi	Risiko	Mitigasi
Jarak dekat, cuaca cerah	Rendah	Jadwal standar
Penerima banyak, waktu siang	Sedang	Mulai lebih awal

Kondisi	Risiko	Mitigasi
Jarak jauh, jalan rusak	Tinggi	Prioritas rute
Hujan lebat, kendaraan kecil	Tinggi	Kendaraan alternatif
Data belum lengkap	Tidak dinilai	Lengkapi dan validasi

Pada kondisi risiko sedang, penambahan buffer waktu dapat dilakukan dengan memajukan waktu keberangkatan atau mengurangi waktu tunggu pada titik sebelumnya. Pada risiko tinggi, mitigasi perlu mempertimbangkan kapasitas kendaraan, urutan tujuan, dan kemungkinan pembagian muatan. Rekomendasi tersebut harus dicatat agar efektivitasnya dapat dievaluasi setelah pengiriman selesai.

Data realisasi yang perlu dicatat meliputi waktu berangkat, waktu tiba, durasi keterlambatan, kendala aktual, dan tindakan mitigasi. Perbandingan antara prediksi dan realisasi memungkinkan pengelola mengetahui apakah rekomendasi berhasil menurunkan risiko. Catatan tersebut juga membantu mengidentifikasi perubahan pola yang belum tercakup pada dataset awal. Penggunaan model tidak menghilangkan kebutuhan komunikasi dengan pengemudi dan lokasi penerima. Informasi penutupan jalan, kerusakan kendaraan, atau perubahan jadwal sekolah mungkin belum tersedia pada saat prediksi dibuat. Oleh karena itu, sistem ditempatkan sebagai alat bantu keputusan yang melengkapi koordinasi operasional

Rancangan Pengujian Sistem

Pengujian sistem dirancang untuk memastikan bahwa fungsi aplikasi berjalan sesuai kebutuhan sebelum digunakan oleh pengelola distribusi. Pengujian fungsional dapat dilakukan menggunakan pendekatan black-box dengan memeriksa hubungan antara data masukan, proses, dan keluaran tanpa menilai kode program secara langsung. Skenario uji yang direncanakan ditunjukkan pada Tabel 11.

Tabel 11. Rancangan Pengujian Fungsional

Fungsi	Skenario	Hasil yang Diharapkan
Login	Akun benar/salah	Masuk atau ditolak
Input	Kolom kosong	Pesan validasi
Prediksi	Data lengkap	Kelas risiko tampil
Riwayat	Simpan hasil	Data tercatat
Laporan	Pilih periode	Laporan terbentuk
Model	Berkas tidak ada	Peringatan sistem

Selain pengujian fungsional, konsistensi model perlu diuji dengan memasukkan data yang sama berulang kali dan memastikan hasilnya tetap sama selama versi model tidak berubah. Pengujian batas dilakukan pada nilai jarak dan jumlah penerima terendah maupun tertinggi. Data kategori yang tidak dikenal harus ditolak atau dipetakan melalui prosedur yang telah ditentukan, bukan diproses secara diam-diam.

Rancangan pada Tabel 11 merupakan rencana pengujian, sehingga status berhasil atau gagal baru dapat diberikan setelah aplikasi benar-benar dijalankan. Bukti pengujian sebaiknya berupa tangkapan layar, waktu

pelaksanaan, data masukan, keluaran aktual, dan catatan koreksi. Pemisahan antara rancangan dan hasil pengujian diperlukan agar laporan penelitian tidak mengklaim pengujian yang belum dilakukan.

Validitas dan Keterbatasan Penelitian

Validitas internal berkaitan dengan kesesuaian antara variabel input, proses pembentukan kelas, dan metode evaluasi. Penggunaan stratified 5-fold cross validation membantu menjaga proporsi kelas pada setiap lipatan, tetapi ukuran dataset yang hanya 100 data tetap membatasi kestabilan estimasi. Perbedaan beberapa data pada setiap lipatan dapat memengaruhi nilai evaluasi secara nyata.

Validitas eksternal berkaitan dengan kemampuan model diterapkan pada kondisi di luar dataset. Dataset penelitian berupa sampel skenario operasional di Kabupaten Soppeng, sehingga hasilnya belum dapat digeneralisasi sebagai kinerja distribusi MBG di seluruh wilayah. Perbedaan karakter jalan, jarak, jumlah penerima, cuaca, dan sistem kerja dapat menghasilkan pola risiko yang berbeda.

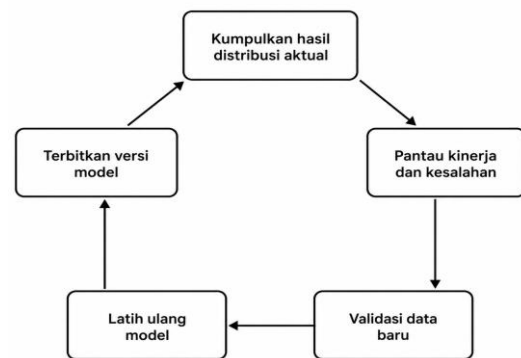
Keterbatasan berikutnya adalah belum tersedianya variabel dinamis seperti kepadatan lalu lintas, waktu antre muat, kerusakan kendaraan, dan perubahan jadwal penerima. Variabel tersebut dapat menjelaskan keterlambatan yang tidak tertangkap oleh enam variabel utama. Penelitian lanjutan perlu merekam waktu rencana dan waktu aktual agar label keterlambatan dibentuk dari selisih waktu yang terukur.

Ketidakeimbangan kelas juga perlu diperhatikan karena kelas Sangat Terlambat hanya terdiri atas 18 data. Akurasi keseluruhan dapat terlihat tinggi meskipun kemampuan model pada kelas minoritas lebih rendah. Oleh sebab itu, evaluasi per kelas, confusion matrix, recall, dan F1-score tetap diperlukan bersama akurasi [11].

Dari sisi etika penggunaan, prediksi tidak boleh menjadi satu-satunya dasar penilaian kinerja petugas atau pengemudi. Kesalahan prediksi dapat terjadi akibat data yang tidak lengkap maupun kejadian di luar kendali. Setiap keputusan yang berdampak pada perubahan rute, jadwal, atau pembagian beban perlu disertai verifikasi manusia dan pencatatan alasan keputusan.

Pemeliharaan Model dan Rencana Implementasi

Model yang telah diintegrasikan ke aplikasi perlu dipantau setelah digunakan. Perubahan pola distribusi dapat terjadi ketika jumlah penerima bertambah, rute berubah, kendaraan diganti, atau musim berganti. Apabila pola data baru berbeda dari data pelatihan, kinerja model dapat menurun meskipun aplikasi tetap berjalan secara teknis.



Gambar 8. Siklus Pemeliharaan Model Prediksi

Siklus pada Gambar 8 dimulai dari pengumpulan realisasi distribusi. Kinerja dipantau dengan membandingkan kelas prediksi dan status aktual, kemudian data baru diperiksa kelengkapan serta konsistensinya. Pelatihan ulang dilakukan apabila jumlah data telah memadai atau terjadi penurunan kinerja yang bermakna. Model baru harus divalidasi sebelum menggantikan versi sebelumnya. Setiap versi model sebaiknya memiliki nomor versi, tanggal pelatihan, jumlah data, daftar variabel, parameter, dan hasil evaluasi. Informasi tersebut membantu proses audit serta memungkinkan pengelola kembali ke versi sebelumnya apabila model baru menimbulkan masalah. Berkas pipeline prapemrosesan harus disimpan bersama model agar proses inferensi tetap konsisten.

Implementasi dapat dilakukan secara bertahap. Tahap awal menggunakan sistem sebagai rekomendasi pendamping tanpa mengubah prosedur distribusi. Tahap berikutnya membandingkan prediksi dengan realisasi selama beberapa periode. Setelah manfaat dan keterbatasannya dipahami, sistem dapat digunakan untuk membantu penyusunan jadwal, prioritas rute, dan evaluasi distribusi secara lebih luas.

Dengan pendekatan bertahap, manfaat sistem dapat diukur melalui perubahan proporsi keterlambatan, ketepatan rekomendasi, waktu respons petugas, dan kemudahan penggunaan. Hasil pengukuran tersebut menjadi dasar untuk memperbaiki antarmuka, menambah variabel, atau menyesuaikan model sebelum penerapan pada cakupan yang lebih besar.

Tata Kelola Data dan Evaluasi Operasional

Keberhasilan sistem prediksi bergantung pada kualitas data yang dimasukkan dan konsistensi pencatatan hasil distribusi. Setiap aktivitas perlu memiliki identitas unik yang menghubungkan tanggal, titik tujuan, kendaraan, petugas, dan waktu pelaksanaan. Identitas tersebut mencegah satu aktivitas tercatat lebih dari sekali serta memudahkan penelusuran ketika terdapat perbedaan antara prediksi dan realisasi.

Data input perlu melalui pemeriksaan kelengkapan, kesesuaian tipe, dan kewajaran nilai. Jarak tidak boleh bernilai negatif, jumlah penerima harus berada pada rentang yang dapat dilayani, dan kategori cuaca maupun kondisi jalan harus mengikuti daftar yang telah ditetapkan. Perubahan istilah kategori harus dicatat karena perbedaan ejaan dapat menghasilkan kolom encoding baru dan mengganggu konsistensi model.

Akses data dibedakan berdasarkan peran pengguna. Administrator mengelola akun, parameter, dan versi model, sedangkan petugas operasional memasukkan data distribusi serta melihat rekomendasi. Pengguna laporan hanya memperoleh akses baca terhadap ringkasan hasil. Pembatasan ini mengurangi risiko perubahan data atau konfigurasi oleh pihak yang tidak memiliki kewenangan.

Tabel 12. Indikator Evaluasi Operasional Sistem

Indikator	Pengukuran dan Tujuan
Ketepatan prediksi	Rasio prediksi benar untuk menilai kinerja model
Recall risiko tinggi	Proporsi risiko tinggi terdeteksi agar kasus tidak terlewat
Waktu respons	Durasi input hingga hasil untuk menilai kecepatan layanan
Kelengkapan data	Proporsi kolom lengkap untuk menilai kualitas pencatatan
Kepatuhan rekomendasi	Tindakan yang diterapkan untuk menilai penggunaan sistem
Perubahan keterlambatan	Perbandingan antarperiode untuk menilai manfaat operasional

Indikator pada Tabel 12 perlu dibaca secara bersama. Akurasi yang tinggi belum tentu memadai apabila model gagal mendeteksi kelas Sangat Terlambat, sehingga recall kelas risiko tinggi menjadi indikator penting. Waktu respons dan kelengkapan data menggambarkan kualitas sistem informasi, sedangkan kepatuhan rekomendasi menunjukkan apakah keluaran model dapat diterapkan oleh petugas. Evaluasi manfaat operasional dilakukan dengan membandingkan periode sebelum dan sesudah penggunaan sistem. Perbandingan perlu mempertimbangkan perubahan jumlah tujuan, musim, kondisi jalan, dan kapasitas kendaraan agar penurunan atau peningkatan keterlambatan tidak langsung dikaitkan dengan model. Catatan tindakan mitigasi membantu menjelaskan hubungan antara rekomendasi dan hasil distribusi.

Pelaporan dapat disusun dalam bentuk ringkasan harian dan rekap bulanan. Ringkasan harian berisi daftar tujuan berisiko, rekomendasi, dan status realisasi. Rekap bulanan memuat distribusi kelas, faktor dominan, kesalahan prediksi, serta tren keterlambatan. Informasi tersebut digunakan untuk evaluasi rute, penyusunan jadwal, dan perencanaan kebutuhan kendaraan. Data yang digunakan untuk pelatihan ulang harus melalui proses verifikasi. Catatan yang tidak lengkap, duplikat, atau memiliki status yang belum terkonfirmasi dipisahkan dari data pelatihan. Dengan prosedur ini, peningkatan jumlah data tidak mengorbankan kualitas model. Dataset, kode pemrosesan, parameter, dan hasil evaluasi juga perlu disimpan agar proses penelitian dan pembaruan model dapat direproduksi.

Pencadangan dilakukan secara berkala pada basis data dan berkas model. Apabila terjadi kegagalan aplikasi, sistem harus dapat dipulihkan tanpa kehilangan riwayat prediksi yang telah tercatat. Prosedur pemulihan, pergantian versi model, dan koreksi data perlu didokumentasikan sebagai bagian dari tata kelola sistem.

Melalui tata kelola tersebut, sistem tidak hanya menghasilkan prediksi, tetapi juga menyediakan dasar evaluasi yang dapat dipertanggungjawabkan. Informasi prediksi, keputusan petugas, dan realisasi pengiriman membentuk siklus pembelajaran yang mendukung perbaikan distribusi secara berkelanjutan.

IV. KESIMPULAN

Penelitian ini menghasilkan model prediksi risiko keterlambatan distribusi pangan Program Makan Bergizi Gratis berbasis machine learning. Dataset sebanyak 100 data sampel skenario operasional digunakan untuk membangun klasifikasi tiga kelas, yaitu Tepat Waktu, Terlambat, dan Sangat Terlambat. Hasil pengujian menunjukkan bahwa Random Forest memperoleh akurasi, presisi, recall, dan F1-score sebesar 0,92. Confusion matrix menunjukkan bahwa model mampu mengenali mayoritas data pada setiap kelas, meskipun masih terdapat kesalahan pada batas kelas Terlambat dan Sangat Terlambat.

Analisis feature importance menunjukkan bahwa jumlah penerima, jarak distribusi, dan waktu distribusi merupakan faktor dominan yang memengaruhi risiko keterlambatan. Sistem prediksi yang diusulkan dapat membantu pengelola MBG dalam mengidentifikasi risiko lebih awal dan menyusun tindakan mitigasi, seperti penjadwalan ulang, penambahan buffer waktu, atau pemilihan kendaraan yang lebih sesuai. Saran untuk penelitian selanjutnya adalah menggunakan data lapangan yang lebih besar dan bersumber dari catatan distribusi aktual, menambahkan variabel waktu tempuh aktual, kepadatan lalu lintas, kapasitas kendaraan, serta menguji algoritma lain seperti Gradient Boosting, XGBoost, atau model berbasis deep learning. Sistem juga perlu diuji langsung pada lingkungan pengguna agar manfaat prediksi dapat diukur secara operasional.

V. REFERENSI

- [1] M. A. Wardana, "Model Explainable Machine Learning untuk Prediksi Kecukupan Gizi Menu MBG," vol. 9, no. April, 2026.
- [2] M. R. Kaylla, H. S. Aulis, Z. Z. Salsabila, P. Studi, M. Logistik, and U. Logistik, "Distribusi Pasokan Bahan Pangan Pada Program Makan Bergizi Gratis (MBG) Daerah Sarijadi," vol. 2, no. 2, pp. 1570–1584, 2026.
- [3] A. E. Serang, A. Sam, L. Bolang, S. L. Hermien, and V. J. Lopian, "JURNAL," vol. 8, no. 6, pp. 1807–1814, 2025.
- [4] A. H. Arayadiba, E. K. Nurnawati, and N. Rahmawati, "Prediksi Tingkat Risiko Keterlambatan Pengiriman Barang Menggunakan Algoritma Decision Tree," vol. 13, no. 01, pp. 51–57, 2025.
- [5] H. K. Pambudi, P. Giri, A. Kusuma, F. Yulianti, and K. A. Julian, "PREDIKSI STATUS PENGIRIMAN BARANG MENGGUNAKAN," vol. 6, no. 2, pp. 100–109, 2020.
- [6] S. P. Sarjana *et al.*, "PERBANDINGAN MODEL MACHINE LEARNING BERBASIS BADIR FRAMEWORK UNTUK DETEKSI FRAUD DAN PREDIKSI," vol. 17, no. 1, pp. 219–231, 2026.